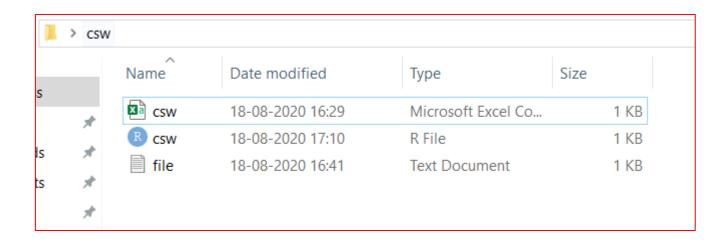
# Custom Stop Words in R

Data Files: (1) file.csv (2) csw.csv

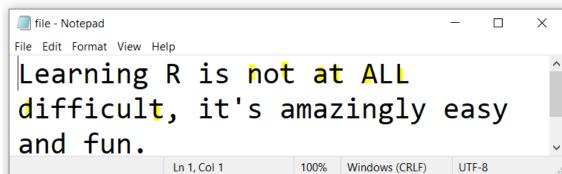


#### Data



4	Α	В
1	not	
2	at	
3	all	
4	difficult	
5		







#### **Set Working Directory**

```
1 # Jesus is Great!
2 setwd("C:/Users/Dr Vinod/Desktop/csw")
3 getwd()
4 library(tm)
```

```
> getwd()
[1] "C:/Users/Dr Vinod/Desktop/csw"
```



#### Read Data

```
6 data <- readLines("file.txt")
7 data</pre>
```

```
> data <- readLines("file.txt")
Warning message:
In readLines("file.txt") : incomplete final line found on 'file.txt'
> data
[1] "Learning R is not at ALL difficult, it's amazingly easy and fun."
```



```
8 text <- VectorSource(data)
9 text</pre>
```

# Put it in a vector



```
text <- VectorSource(data)</pre>
  text
$`encoding
[1]
$1ength
[1] 1
$position
[<u>1</u>] 0
$reader
function (elem, language, id)
    if (!is.null(elem$uri))
        id <- basename(elem$uri)</pre>
    PlainTextDocument(elem$content, id = id, language = language)
<bytecode: 0x00000000c427950>
<environment: namespace:tm>
$content
[1] "Learning R is not at ALL difficult, it's amazingly easy and fun."
attr(,"class")
[1] "VectorSource" "SimpleSource" "Source"
```

#### **Create Corpus**

```
10 text1 <- VCorpus(text)
11 text1</pre>
```

```
> text1 <- VCorpus(text)
> text1
<<VCorpus>>
Metadata: corpus specific: 0, document level (indexed): 0
Content: documents: 1
```



### Import csv file of custom stop words



```
12 #_____custom stop words
13 csw <- read.csv("csw.csv", header = FALSE)
14 csw # as data frame
15 class(csw) # data frame
16 csw1 <- as.character(csw$V1)
17 csw1 # char (in a row)
18 class(csw1) # char</pre>
```

```
csw <- read.csv("csw.csv", header = FALSE)</pre>
> csw # as data frame
         V1
        not
         at
        a11
4 difficult
> class(csw) # data frame
[1] "data.frame"
> csw1 <- as.character(csw$V1)</pre>
> csw1 # char (in a row)
[1] "not"
                 "at"
                              "all"
                                           "difficult"
> class(csw1) # char
[1] "character"
```

# Adding my stop words with default set

```
19 mySW <- c(csw1, stopwords())
20 mySW # first 4 are custom words</pre>
```

```
mySW <- c(csw1, stopwords())</pre>
        first 4 are custom words
                                                               44.0
                    "at"
                                                 "difficult"
     "not"
                                  "a]]"
     "me"
                    "my"
                                                 "we"
                                                               "our"
                                  "myself"
     "ours"
                    "ourselves"
                                  "you"
                                                 "your"
                                                               "yours"
                                                 "him"
                                                               "his"
[16] "yourself"
                    "yourselves"
                                  "he"
```



```
"each"
                                                  "more"
                                                                 "most"
[161]
      "both"
                                    "few"
      "other"
                                                  "no"
                                                                 "nor"
                     "some"
                                    "such"
[166]
                                                                 "so"
[171]
      "not"
                     "only"
                                    "own"
                                                  "same"
[176] "than"
                     "too"
                                    "very"
```

#### Observe L, R & ALL $\rightarrow$ I, r as all

```
21 #
22 text2 <- tm_map(text1, content_transformer(tolower))
23 text2[[1]]$content
24</pre>
```

```
> text2 <- tm_map(text1, content_transformer(tolower))
> text2[[1]]$content
[1] "learning r is not at all difficult, it's amazingly easy and fun."
```



## Stop Words, removed!

```
24
25 text3 <- tm_map(text2, removeWords, mySW)
26 text3[[1]]$content
27</pre>
```

```
> text3 <- tm_map(text2, removeWords, mySW)
> text3[[1]]$content
[1] "learning r amazingly easy fun."
```





### You got what you wanted!

```
27
28 text4 <- tm_map(text3, stripWhitespace)
29 text4[[1]]$content</pre>
```

```
> text4 <- tm_map(text3, stripWhitespace)
> text4[[1]]$content
[1] "learning r , amazingly easy fun."
```

